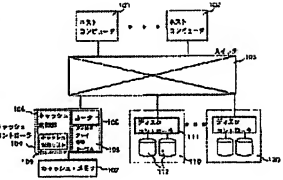
DATA STORAGE SYSTEM

Publication number: JP9198308 (A)
Publication date: 1997-07-31
Inventor(s): FUJIBAYASHI AKIRA; TAKAMOTO YOSHIFUMI
Applicant(s): HITACHI LTD
Classification:
- International: *G06F12/08; G06F3/06;* G06F12/08; G06F3/06; (IPC1-7): G06F12/08; G06F3/06; G06F12/08
- European:
Application number: JP19960023202 19960117
Priority number(s): JP19960023202 19960117

Also published as:
JP3776496 (B2)

Abstract of **JP 9198308 (A)**

PROBLEM TO BE SOLVED: To effectively use a cache memory and to improve data access performance in a data storage system using a switch. SOLUTION: Plural host computers 101 and 102 are connected with secondary storage devices 110 and 120 through the switch 103. A cache memory 107 provided with a cache controller 104 common to the secondary storage devices 110 and 120 is switch-connected to the secondary storage devices in parallel. At the time of accessing data, the host computers 101 and 102 transmit data to the cache memory 107 through the switch. When a cache error occurs in the cache memory 107, a disk array management table 108 is referred to, a port to which the secondary storage device corresponding to a logic volume in a packet is connected is specified and the secondary storage device is accessed.; Data transmitted from the secondary storage device is stored in the cache memory 107 and data is transferred to the host computers 101 and 102 through the switch 103.



(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平9－198308

(43)公開日　平成９年(1997)７月31日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ		技術表示箇所
Ｇ０６Ｆ　12/08		7623－5Ｂ	Ｇ０６Ｆ　12/08	Ｈ	
		7623－5Ｂ		Ｐ	
	３２０	7623－5Ｂ		３２０	
3/06	３０１		3/06	３０１Ｍ	
	３０２			３０２Ａ	

審査請求　未請求　請求項の数５　ＦＤ　（全 10 頁）

(21)出願番号　特願平８－23202

(22)出願日　平成８年(1996)１月17日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

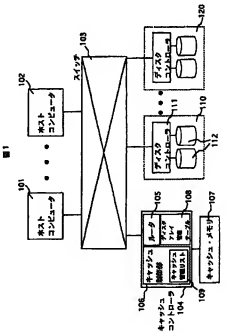
(72)発明者　藤林　昭
東京都国分寺市東恋ケ窪１丁目280番地
株式会社日立製作所中央研究所内

(72)発明者　高本　良史
東京都国分寺市東恋ケ窪１丁目280番地
株式会社日立製作所中央研究所内

(74)代理人　弁理士　筒井　虎　（外１名）

(54)【発明の名称】　データ記憶システム

(57)【要約】
【課題】　スイッチを用いたデータ記憶システムにおいて、キャッシュメモリを有効に活用してデータアクセス性能を向上することにある。
【解決手段】　複数のホストコンピュータ101,102と２次記憶装置110,120はスイッチ103を介して接続され、２次記憶装置に共通のキャッシュコントローラ104を備えるキャッシュメモリ107が２次記憶装置と並列にスイッチに接続される。ホストコンピュータはデータをアクセスするときパケットをスイッチを介してキャッシュメモリに送出する。キャッシュメモリでキャッシュミスが生じたときは、ディスクアレイ管理テーブル107を参照してパケット内の論理ボリュームに対応する２次記憶装置が接続されているポートを特定して２次記憶装置をアクセスし、２次記憶装置から送られてくるデータをキャッシュメモリに格納し、データをスイッチを介してホストコンピュータに転送する。



1

【特許請求の範囲】

【請求項１】　複数のホストコンピュータと、複数の２次記憶装置とから構成され、該複数のホストコンピュータと該複数の２次記憶装置との間をスイッチにより接続するデータ記憶システムにおいて、該スイッチに該複数の２次記憶装置と並列に独立したキャッシュメモリを接続し、該キャッシュメモリは該複数の２次記憶装置の該スイッチを介したキャッシュメモリであることを特徴とするデータ記憶システム。

【請求項２】　請求項１記載のデータ記憶システムにおいて、

前記キャッシュメモリは、前記ホストコンピュータのデータ転送要求に対して、要求されたデータを検索し、検索の結果該データがキャッシュメモリ内にない場合には、前記スイッチを介して該データが格納されている２次記憶装置から該データを取り込み、該キャッシュメモリに格納した後にホストコンピュータに該データを転送することを特徴とするデータ記憶システム。

【請求項３】　請求項２記載のデータ記憶システムにおいて、

前記キャッシュメモリはキャッシュコントローラを備え、該キャッシュコントローラは、複数の各２次記憶装置が接続される前記スイッチのポート番号とホストコンピュータの認識している論理ボリュームを対応させたディスクアレイ管理テーブルを有し、前記検索の結果前記データがキャッシュメモリ内にない場合、該ディスクアレイ管理テーブルを参照して２次記憶装置をアクセスすることを特徴とするデータ記憶システム。

【請求項４】　請求項３記載のデータ記憶システムにおいて、

前記各ホストコンピュータのオペレーションシステムに前記ディスクアレイ管理テーブルを設け、該各ホストコンピュータは、前記複数の２次記憶装置のいずれかに直接アクセスするとき、該ディスクアレイ管理テーブルを参照して２次記憶装置をアクセスすることを特徴とするデータ記憶システム。

【請求項５】　請求項３記載のデータ記憶システムにおいて、

前記スイッチは、前記各ホストコンピュータが前記キャッシュメモリを使用するか直接前記２次記憶装置を使用するかを示すテーブルと直接前記２次記憶装置を使用する場合に用いられる前記ディスクアレイ管理テーブルと同様のテーブルからなるホスト管理テーブルを有し、前記ホストコンピュータからのアクセス要求に応じて前記ホスト管理テーブルを参照して前記キャッシュメモリまたは前記２次記憶装置で前記ホストコンピュータからのアクセスデータを転送することを特徴とするデータ記憶システム。

【発明の詳細な説明】

【０００１】

2

【発明の属する技術分野】本発明は、複数のホストコンピュータと複数の２次記憶装置をスイッチを利用して接続する構成のデータ記憶システムに関する。

【０００２】

【従来の技術】一般的なデータ処理システムは、ホストコンピュータと２次記憶装置から構成されている。２次記憶装置として使用されるのは主に磁気ディスク装置である。ここで発明者のいう磁気ディスク装置は単体のディスクドライブまたはディスクアレイを意味する。本発明の利用分野として挙げたスイッチを利用したデータ記憶システムの一例が特開平７－４４３２２号において記述されている。

【０００３】スイッチを用いてホストコンピュータと磁気ディスク装置を接続する構成のデータ記憶システムでは、ホストコンピュータおよび磁気ディスク装置の台数の変更にも柔軟に対応することが可能でスケーラビリティに優れる。このスイッチを利用したホストコンピュータと磁気ディスク装置間のデータ転送はパケット交換によって実現される。これを簡単に説明すると、ホストコンピュータはコマンドパケット（例えば、ポート番号１番；セクタ番号２；データサイズ５１２Ｂyte；Ｒeadのような構成）を送信し、そのコマンドパケットをスイッチが解析してホストコンピュータと要求されたポート間の接続が確立したのちに、ホストコンピュータとそのポートに接続している磁気ディスク装置間でデータ転送が開始すると言う手順になる。また、従来のバススイッチと異なり、ある１組のホストコンピュータとの磁気ディスク装置の間でデータ転送が行われていても、データバスは占有されず、同時に他のホストコンピュータと磁気ディスク装置の間でもデータ転送が行える。ここで、データの読出しに関して、磁気ディスク装置では、ホストコンピュータから頻繁にアクセス要求のあるデータに対して、その都度読出し動作を行うと、メカニカルな動作を伴うためデータ転送に時間がかかるという問題がある。この問題を解決するために、磁気ディスク装置は通常キャッシュメモリを備えており、頻繁にアクセス要求のあるデータは、そのキャッシュメモリに格納することでそのデータの読出しに対するメカニカルな動作を省略し性能の向上を図っている。しかし、従来例の一つとして挙げたシステム構成では、従来のバススイッチを本発明の用いているスイッチに置き換え、ホストコンピュータとのインターフェース以下の部分はアレイコントローラが磁気ディスク装置とホストコンピュータ間のデータ転送を制御している従来の一般的構成である。この様な構成の場合には、接続された磁気ディスク装置の各々にキャッシュメモリが分散されて配置されており、それぞれのキャッシュが有効に利用されていない。

【０００４】

【発明が解決しようとする課題】従来例として挙げたデータ記憶システムおいて、接続している複数の磁気ディ

スク装置のそれぞれにキャッシュメモリが備わっている場合、性能向上の為のキャッシュメモリが、接続している磁気ディスク装置それぞれに分散して配置されることになり、それぞれのキャッシュを有効に利用できない。例えば、ある一つの磁気ディスク装置に格納されているデータに頻繁にアクセス要求が送られてくるような場合には、他の磁気ディスク装置に備わっているキャッシュメモリはほとんど利用されないことになり無駄になってしまう。

【０００５】本発明の目的は、スイッチを用いたデータ記憶システムにおいて、キャッシュメモリを有効に活用してデータアクセス性能を向上することにある。

【０００６】

【課題を解決するための手段】上記目的を達成するため、本発明は、複数のホストコンピュータと、複数の２次記憶装置とから構成され、該複数のホストコンピュータと該複数の２次記憶装置との間をスイッチにより接続するデータ記憶システムにおいて、該スイッチに該複数の２次記憶装置と並列に独立したキャッシュメモリを接続し、該キャッシュメモリは該複数の２次記憶装置の該スイッチを介したキャッシュメモリであるようにしている。また、前記キャッシュメモリは、前記ホストコンピュータのデータ転送要求に対して、要求されたデータを検索し、検索の結果該データがキャッシュメモリ内にない場合には、前記スイッチを介して該データが格納されている２次記憶装置から該データを取り込み、該キャッシュメモリに格納した後にホストコンピュータに該データを転送するようにしている。前記キャッシュメモリはキャッシュコントローラを備え、該キャッシュコントローラは、複数の各２次記憶装置が接続される前記スイッチのポート番号とホストコンピュータの認識している論理ボリュームを対応させたディスクアレイ管理テーブルを有し、前記検索の結果前記データがキャッシュメモリ内にない場合、該ディスクアレイ管理テーブルを参照して２次記憶装置をアクセスするようにしている。また、前記各ホストコンピュータのオペレーションシステムに前記ディスクアレイ管理テーブルを設け、該各ホストコンピュータは、前記複数の２次記憶装置のいずれかに直接アクセスするとき、該ディスクアレイ管理テーブルを参照して２次記憶装置をアクセスするようにしている。また、前記スイッチは、前記各ホストコンピュータが前記キャッシュメモリを使用するか直接前記２次記憶装置を使用するかを示すテーブルと直接前記２次記憶装置を使用する場合に用いられる前記ディスクアレイ管理テーブルと同様のテーブルからなるホスト管理テーブルを有し、前記ホストコンピュータからのアクセス要求に応じて前記ホスト管理テーブルを参照して前記キャッシュメモリまたは前記２次記憶装置に前記ホストコンピュータからのアクセスデータを転送するようにしている。

【０００７】

【実施例】本発明の提供するデータ記憶システムを以下に図面を示し実施例を参照して詳細に説明する。図１は、本発明によるデータ記憶システムの構成をブロック図で示したものである。１０１、１０２はホストコンピュータであり、１０３はスイッチであり、１０４はキャッシュコントローラであり、ルータ１０５とキャッシュ制御部１０６から構成される。１０７はキャッシュメモリである。１０８はキャッシュコントローラ１０４の持つディスクアレイ管理テーブルである。１０９はキャッシュ制御部１０６がキャッシュ検索時に参照するキャッシュ管理リストである。１１０、１２０は磁気ディスク装置であり、この図１中では一例としてディスクアレイ装置としており、ディスクコントローラ１１１、複数のディスク装置１１２から構成されている。磁気ディスク装置１２０も同様の構成である。この図に示すように本発明はキャッシュメモリ１０７を、複数の磁気ディスク装置１１０、１２０と共にスイッチ１０３に並列に接続することで、キャッシュメモリ１０７は前記磁気ディスク装置１１０、１２０及びホストコンピュータ１０１、１０２と、同様の１つのデバイスのごとくスイッチにより容易にアクセスできる構成となる。そして、磁気ディスク装置１１０、１２０個々にはキャッシュメモリを設けず、キャッシュメモリ１０７を磁気ディスク装置１１０、１２０が共用する。

【０００８】ここで、スイッチ１０３の内部構造の一例を図７に示す。７０１はデータのシリアル／パラレル変換を行う部分（Ｓ／Ｐ）である。７０２はスイッチ制御装置である。７０３はスイッチ機構である。ここで、スイッチ１０３の動作の一例を簡単に説明する。図７の中に示すような構成のパケット７０４がポート１に接続されているホストコンピュータから発行される。パケットは送信先ポート番号とデータ部からなる。データ部の内容は７０５に示すように、コマンド、論理ボリューム（Ｌｖｏｌ＃）、ブロック番号（ＢＬＫ＃）、ホスト番号（Ｈｏｓｔ＃）からなる。送信されてきたパケットの送信先ポートをみて、スイッチ制御装置７０２がそのポートとの接続を確立するように制御線を通じて信号を送出する。送信先ポートとの接続が確立したのち、送信先に向けてパケットまたはデータが送られる。

【０００９】次に上記システム構成においてキャッシュメモリ１０７を有効利用するために用いるデータ転送方法について述べる。図１のキャッシュ制御部１０６で行うホストコンピュータからのデータ読出し要求に対するデータ転送のフローチャートを図２に示す。先ず、ホストコンピュータはデータ読出し要求をキャッシュメモリの接続ポートに対して送る。ステップ２０１において、要求されたデータをキャッシュ制御部１０６において、送信されてきたパケット内の論理ボリューム（Ｌｖｏｌ）とブロック番号ＢＬＫに基づいてキャッシュ管理リスト１０９を参照しキャッシュメモリ１０７内にデータが存在す

るかどうか検索する。次に判断ステップ２０２において、要求されたデータがキャッシュメモリ１０７内に存在する（キャッシュＨｉｔ）場合には、ステップ２０５において、要求されたデータをホストコンピュータに転送する。要求されたデータが存在しない（キャッシュＭｉｓｓ）場合には、ステップ２０３において、ルータに制御を移す。続いてステップ２０４においては、ルータの制御で磁気ディスク装置から送られて来たデータをキャッシュ制御部１０６がキャッシュメモリ１０７に格納し、キャッシュ管理リストを更新する。この時、キャッシュメモリの容量一杯までデータがすでに格納されている場合は、データの追い出しが必要となる。現在追い出しの手法は様々な方法が利用されているが、ここでは最も使用頻度が少なく最も古いデータを追い出しの候補に選ぶ方法を使用することとするが、その他の方法用いたとしても本発明の実施にはなんら問題はない。そして、ステップ２０５において、ホストコンピュータに要求されたデータを転送する。

【００１０】ここで、図２のステップ２０１においてキャッシュ制御部が参照するキャッシュ管理リストの概略を図９に示す。キャッシュコントローラ内のルータが受信したパケット内の論理ボリューム番号とブロック番号がキャッシュ制御部に渡される。キャッシュメモリ内のデータはキャッシュ管理リストによりブロック単位に管理されている。現在使用中（キャッシュメモリ内にデータが存在する）論理ボリューム番号とブロック番号のリストと未使用のリストを持っており、渡された論理ボリューム番号とブロック番号が使用リスト中に存在するかどうか検索する。

【００１１】図３は前記ルータに制御を移した後の処理をフローチャートで示したものである。ここで、ルータの機能は、簡略に説明すると、ホストコンピュータから転送されたパケット内のディスクコマンドを解析し、その結果に基づいて所定の磁気ディスク装置を選択しコマンドやデータをルーティングすることである。図３中のステップ３０１において、ルータは送信されてきたパケット内の論理ボリュームをみて、ディスクアレイ管理テーブル（以下、図４において説明する。）を参照し、その論理ボリュームに対応する磁気ディスク装置が接続されているポートを特定する。ステップ３０２において、特定したポートに対しパケットを送出する。ステップ３０３において、このパケットを受けた磁気ディスク装置から転送されてくるデータを受け取る。

【００１２】これまで述べてきた実施例の通り、ホストコンピュータへのデータ転送時は常時キャッシュメモリ１０７を使用することで、例えば磁気ディスク装置１０１に格納されているデータに頻繁にアクセスがあるような場合でも、他の磁気ディスク装置に格納されているデータもキャッシュメモリ１０７に格納しておくことができ、すべての磁気ディスク装置に対してキャッシュメモリ１０７が有効利用される。

【００１３】図４は前記ディスクアレイ管理テーブルである。ホストコンピュータのオペレーティングシステム（以下、ＯＳと記述する）は、この論理ボリューム（Ｌｖｏ１）とブロック（ＢＬＫ）によりデータを指定してアクセス要求を発行する。実際のデータは複数ある磁気ディスク装置内に格納されているので、その磁気ディスク装置の接続されているポートとの対応を取るためにこのディスクアレイ管理テーブルを用いる。このディスクアレイ管理テーブルをキャッシュコントローラ１０４が持ち、前記実施例のようにホストコンピュータへのデータ転送時に常時キャッシュメモリ１０７を利用することで、ホストコンピュータのＯＳは論理ボリュームを用いてデータアクセス要求を発行することになり、複数の磁気ディスク装置が接続されていることとその内部のデータ配列の構成をユーザーに意識させず、単一の磁気ディスク装置の使用環境を提供する。

【００１４】図５では本発明の提供するシステムにおいて、実際のデータ転送の一例を簡略に示している。ここでは、ホストコンピュータ１０１からデータ１への読出し要求が発行され、ホストコンピュータ１０２からデータ２への読出し要求が発行されている場合を考える。データ１は前記キャッシュメモリ内に現在格納されており、データ２はこのキャッシュメモリ内に無く磁気ディスク装置１２０に格納されている。ホストコンピュータ１０１からのデータ１の読出し要求を図５中に示すような構成のパケット５０１として送信する。図２で示したフローチャートに従い、先ず最初に上記キャッシュメモリ１０７内が検索される。データ１はキャッシュＨｉｔするので、キャッシュメモリ１０７からホストコンピュータ１０１に転送される。同様にホストコンピュータ１０２はパケット５０２を送信して、上記キャッシュメモリ１０７内が検索されるがキャッシュＭｉｓｓとなり、キャッシュコントローラ内のルータが、図４で説明したディスクアレイ管理テーブルを参照してデータ２の格納位置を磁気ディスク装置１２０の接続されているポートと特定し、これに対してデータ２の読出し要求のパケット５０３を発行して、データを取込み、上記キャッシュメモリ１０７内に格納してから、ホストコンピュータ１０２に転送する。

【００１５】これまでに説明した実施例では接続するすべてのホストコンピュータは本発明の提供するキャッシュメモリにアクセスするシステムになっている。しかし、ホストコンピュータによっては、使用するアプリケーションの性質により、本発明の提供するキャッシュメモリより直接磁気ディスク装置にアクセスする方が性能が良いという場合がある。これに対して、以下に説明する実施例で対応する。一つの実施例は、ディスクアレイ管理テーブルをホストコンピュータのＯＳにも持たせることである。これにより、ホストコンピュータには論理

ボリュームと実際にデータの格納されている磁気ディスク装置が接続されているポートの対応が分かっているので、発行するパケットの送信先ポートにキャッシュメモリまたは磁気ディスク装置の接続されているポートを指定することで、本発明のキャッシュメモリの使用、不使用が選択可能になる（概略図を図６に示す。）

もう一つの実施例は前記スイッチにホストコンピュータのアクセス要求の管理を行う構成管理テーブルを持たせることである。図８にその概略図を示す。このホスト管理テーブル８０１は、スイッチ１０３内のスイッチ制御装置７０２に接続され、送信されてきたパケット内のホスト番号とホスト管理テーブル８０１を照合して、キャッシュメモリ１０７を使用するホストコンピュータと使用しないホストコンピュータを判断し、ホストコンピュータの発行する転送要求先をキャッシュメモリ１０７か、要求するデータの格納されている磁気ディスク装置かに振り分ける。例えば、ホストコンピュータの発行するアクセス要求をホスト管理テーブル８０１を参照してキャッシュメモリを使用しないホストである場合、ホスト管理テーブルはディスクアレイ管理テーブルと同様のテーブルを持っており、これを参照して接続先ポートを、指定された論理ボリュームに対応する磁気ディスク装置が接続しているポートに指定する。上記２通りの実施例により、本発明の提供するキャッシュメモリの使用・不使用の選択性を有するシステムを実現できる。なお、上記実施例では、ホストコンピュータ、磁気ディスク装置、キャッシュメモリの接続のためにスイッチによる接続方式を用いているが、スイッチによる接続方式に代えてループ状の接続方式を用いることも可能である。

【００１６】

【発明の効果】本発明により、スイッチを用いたデータ記憶システムにおいて、キャッシュメモリを有効に活用してデータアクセス性能が向上する。さらに、ホストコンピュータの使用するアプリケーションによってキャッシュメモリを使用するか、キャッシュメモリを不使用として直接磁気ディスク装置にアクセスするかの選択性を有する効率のよいデータ記憶システムを可能とする。

【図面の簡単な説明】

【図１】本発明によるキャッシュメモリを備えるスイッチを利用したデータ記憶システムの構成を示すブロック図である。

【図２】ホストコンピュータからの読出し要求に対するキャッシュ制御部の処理のフローチャートを示す図である。

【図３】ルータがデータをキャッシュメモリに格納する場合の処理のフローチャートを示す図である。

【図４】ディスクアレイ管理テーブルの一例を示す図である。

【図５】本発明における実際のデータの流れの一例を示した概略図である。

【図６】ディスクアレイ管理テーブルをホストコンピュータのＯＳに持たせた場合の本発明のデータ記憶システムの構成を示すブロック図である。

【図７】スイッチの構造の概略図である。

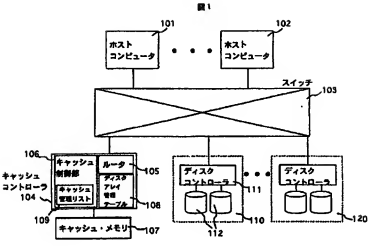
【図８】スイッチがホスト管理テーブルを備える場合の本発明のデータ記憶システムの構成を示すブロック図である。

【図９】キャッシュ管理リストの概略図である。

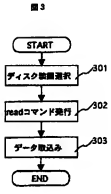
【符号の説明】

１０１、１０２　ホストコンピュータ
１０３　スイッチ
１０４　キャッシュコントローラ
１０５　ルータ
１０６　キャッシュ制御部
１０７　キャッシュメモリ
１０８　ディスクアレイ管理テーブル
１１０、１２０　磁気ディスク装置
１１１　ディスクコントローラ
１１２　ディスク装置
７０１　Ｓ／Ｐ
７０２　スイッチ制御装置
７０３　スイッチ機構
８０１　ホスト管理テーブル

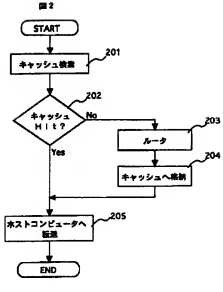
【図１】



【図３】



【図２】



【図４】

図４

論理ボリューム番号 (401)	接続先ポート番号 (402)
0	ポート1
1	ポート1
2	ポート2
3	ポート2
4	ポート3
⋮	⋮

ディスクアレイ管理テーブル

【図5】

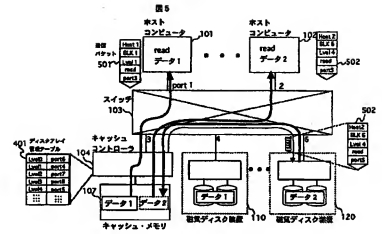


図5
ホスト
コンピュータ
101
ホスト
コンピュータ
102
read
データ1
read
データ2
port1
port2
スイッチ
103
キャッシュ
コントローラ
104
ディスクアレイ
管理テーブル
401
データ1
データ2
キャッシュ・メモリ
107
データ1
データ2
磁気ディスク装置
110
磁気ディスク装置
120

【図6】

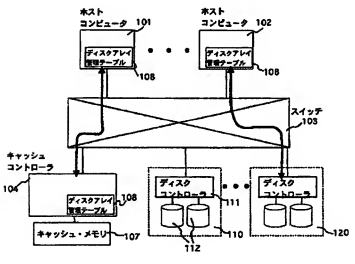
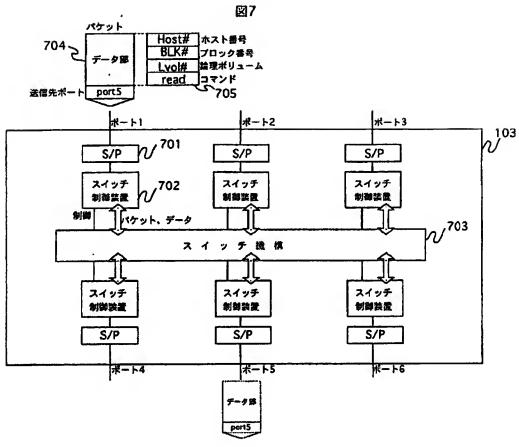


図6
ホスト
コンピュータ
101
ホスト
コンピュータ
102
ディスクアレイ
管理テーブル
108
スイッチ
103
キャッシュ
コントローラ
104
ディスク
コントローラ
111
ディスク
コントローラ
キャッシュ・メモリ
107
110
112
120

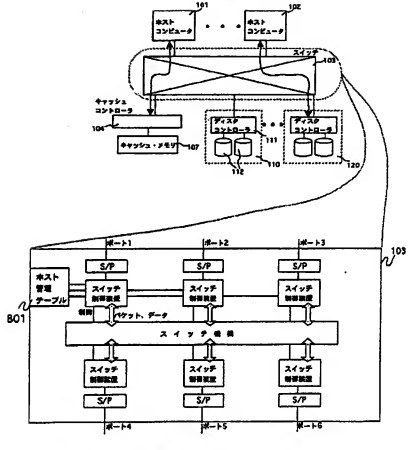
【図7】

図7


パケット
704
データ部
Host#
BLK#
Lvol#
read
ホスト番号
ブロック番号
論理ボリューム
コマンド
705
送信先ポート
port5
ポート1
ポート2
ポート3
103
S/P
701
スイッチ制御装置
702
制御
パケット、データ
703
スイッチ機構
ポート4
ポート5
ポート6
データ部
port5

【図８】

図8

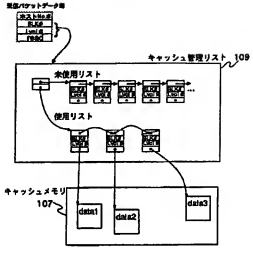


ホスト管理テーブル

host0	use	Lvol0	port4
host1	no use	Lvol1	port4
host2	use	Lvol2	port5
host3	no use	Lvol3	port5
host4	use	Lvol4	port6
host5	use	Lvol5	port6
⋮	⋮	⋮	⋮

【図9】

図9


キャッシュ管理リスト
109
未使用リスト
使用リスト
キャッシュメモリ
107
data1
data2
data3